1989

Relative compromise of statistical databases

M Miller

Jennifer Seberry
University of Wollongong, jennie@uow.edu.au

Relative compromise of statistical databases

Abstract

Statistical databases are databases in which only statistical type of queries are allowed. The results of the statistical queries are intended for statistical use only. However, it has been shown that using only statistical queries it is often possible to infer an individuals's value of a protected field (e.g, using various types of trackers). In such a case we say that the database has been (positively) compromised. Various types of compromise have been studied but until now attention has centred on the inference of exact information from permitted queries. In this paper we introduce a new type of compromise, the 'relative' compromise: a set of records is relatively compromised with respect to a field X if the relative order of magnitude of the X-values of the set is known. This paper shows that even when exact information is protected, relative information may be accessible. We consider several sets of conditions under which this compromise can occur using SUM type of queries of fixed query set size, as well as some of the possible consequences of relative compromise.

Disciplines

Physical Sciences and Mathematics

Publication Details

Miller, M and Seberry, J, Relative compromise of statistical databases, ACSC12 and The Australian Computer Journal, 21(2), 1989, 56-61.

Relative compromise of statistical databases

M. Miller
Department of Mathematics, Statistics and Computing Science, University of New England, Armidale, NSW
J. Seberry
Department of Computer Science, University College, ADFA, Canberra, ACT

Statistical databases are databases in which only statistical type of queries are allowed. The results of the statistical queries are intended for statistical use only. However, it has been shown that using only statistical queries it is often possible to infer an individual's value of a protected field (e.g., using various types of trackers). In such a case we say that the database has been (positively) compromised. Various types of compromise have been studied but until now attention has centred on the inference of exact *information from permitted queries. In this paper we introduce a new type of compromise, the 'relative' compromise: a set of records is relatively compromised with respect to a field X if the relative order of magnitude of the X-values of the set is known. This paper shows that even when exact information is protected, relative information may be accessible. We consider several sets of conditions under which this compromise can occur using SUM type of queries of fixed query set size, as well as some of the possible consequences of relative compromise.*

Keywords and Phrases: Database inference, statistical database, SUM queries, compromise of statistical databases, relative compromise.
CR Categories: H.3.3, H.3.5.

This paper was presented at the Twelfth Australian Computer Science Conference at the University of Wollongong, NSW, 8-10 February 1989.

INTRODUCTION

A database D is a finite set of N records (or tuples) in which each record has a finite number of fields (or attributes). For the purpose of this study we shall assume that one of these fields or attributes, say X, is to be kept secret for all records i in the database. Furthermore, we assume that the elements $x_i \in X$ are real numbers.

A statistical database is a database in which only statistical types of queries are allowed, such as COUNT, SUM, AVG, MIN, MAX. Such a query $q(S; X)$ operates on the values of the attribute X of a subset S (called the query set) of the database. The query set is selected by a characteristic formula. For a more detailed explanation of the terms used here refer to Denning (1982), Chapter 6. We shall not exclude the possibility of using key values in the characteristic formula. For simplicity we identify both a characteristic formula and its corresponding query set by the same symbol ; and we shall write $q(S)$ instead of $q(S; X)$ when X is understood.

A statistical database is to be used for statistical purposes only and the X-values of the individual records are to be protected from disclosure. If a disclosure of a X-value of any of the individual records occurs we say that the database has been compromised. Various types of compromise have been defined (Davida, et al., 1978; Denning, 1982; Dobkin, et al., 1979), for example:

Definition 1: A database is said to be *positively compromised*, or simply *compromised*, if one or more individuals can have their X-values associated with them.

Definition 2: A database is *negatively compromised* if it is known that some particular value is not the X-value of a particular individual.

Definition 3: If all individual records in a subset S of a database D can be compromised we say that the subset S is *completely compromised*.

In the absence of any restrictions on the queries a statistical database can be compromised simply by making the query SUM (S) where S is a characteristic formula that uniquely identifies some individual i. Alternatively, we could deduce the value of x_i from SUM (D) — SUM (D — S). It is therefore obvious that to protect the field X we need to place some restrictions on the allowed queries. To prevent the above compromise we restrict the query set size $k = |S|$ to be within the range $[2, N-1]$, and taking into account possible supplementary knowledge of say $n-1$ individuals, we further restrict k to

$$n \leq k \leq N - n \qquad (*)$$

Since this restriction is necessary (although not sufficient) to prevent compromise we shall assume it from now on.

Relative Compromise

It was shown (Denning, 1982) that even for k restricted to being close to the value of $\frac{N}{2}$ it is possible to compromise a database using the techniques of individual, general, dou-

ble and union trackers. These techniques rest on:

(a) the ability to ask queries whose query sets overlap and/or on
(b) the ability to ask queries of variable query set size (within our general constraint (*)) and/or on
(c) the ability to ask the number of queries needed to make the inferences.

(a) We say that the overlap of a set of queries is λ if any two queries of the set have at most λ individuals in common and some two queries have exactly λ individuals in common. If we restrict the overlap λ to $\lambda = 0$ then compromise using trackers is not possible. However, such a database would not be very useful.

(b) On the other hand, if query set size k were constant then compromise is still possible (by exploiting the overlap for MIN or MAX queries with cleverly chosen query sets; or by solving a system of linear equations for SUM or AVG queries) even with the smallest possible overlap $\lambda = 1$ as long as there is no restriction on the number of queries allowed concerning the individuals involved in the queries[1].

(c) Thus another possible approach is to restrict the number of queries allowed. Using this method it was found that (positive) compromise can be prevented. In particular, Davida et al. (1978) found that it is not possible to compromise a database by asking M queries, each of size k, concerning N individuals, if no two queries overlap in more than one position and if

$$k \leq \frac{\frac{N}{M} + \sqrt{(\frac{N}{M})^2 + 4(N - \frac{N}{M})}}{2}$$

On the other hand, Dobkin, Jones and Lipton (1979) studied the function $M = S(N,k,\lambda,l)$, where M is the smallest number of SUM queries that suffices to compromise the database, k is the query set size (fixed), λ is the query set overlap, and l is the number of X-values known a priori to the user. They found that:

(a) $S(N,k,1,0) \leq 2k - 1$, $N \geq k^2 - k + 1$
(b) $S(N,k,1,1) \leq 2k - 2$, $N \geq (k-1)^2 + 2$
(c) $S(N,k,\lambda+\alpha,\lambda,2\alpha-1) \leq 2k$, $N \geq k^2\lambda + 2\alpha$
(d) $S(N',k\lambda,\lambda,\lambda-1) \leq S(N,k,1,0)$, $N' \geq \lambda k^2$

They also showed that compromise is impossible if

$$N < \frac{k^2-1}{2\lambda} + \frac{k+1}{2}$$

That is, in this case $S(N,k,\lambda,0) = \infty$.

Taking into account possible supplementary knowledge of l individuals, then Dobkin, Jones and Lipton (1979) showed that $S(N,k,\lambda,l) = \infty$, that is, compromise is not possible if

$$N < \frac{k^2-(l+1)^2}{2\lambda} + \frac{k+l+1}{2}$$

In this paper we consider the restriction on the number of queries for SUM type of queries. The compromise of a database using only SUM queries of fixed query set size k always involves the construction of some l linearly independent queries concerning l individuals. Such a set of queries can be expressed as a system of l linearly independent equations in l unknowns. Solving these equations leads to a complete compromise of the l individuals. To prevent this compromise we could restrict the number of queries allowed so that in any system of equations derived from the queries there would always be at least one more unknown than the number of equations. This control could be implemented either in a static or a dynamic way. We shall now consider such a system where we allow SUM queries as long as no subset of all queries asked could be written as a system of l linearly independent equations on l unknowns. We shall show that even this restriction does not prevent a new type of compromise, the "relative compromise".

Definition 4: A subset S ($|S| > 1$) of a database is *relatively compromised* if the relative order of magnitude of the individuals in the subset is known.

Theorem 1: Let S be a subset of a database, $|S| = k + 1$. Then a subset of k individuals of S can be relatively compromised using only k SUM queries with fixed query set size k.

Proof: Construct queries that can be written as the following system of k equations in $k + 1$ unknowns.

$$x_2 + x_3 + \ldots + x_k + x_{k+1} = q_1$$
$$x_1 + x_3 + \ldots + x_k + x_{k+1} = q_2$$
$$\ldots$$
$$x_1 + x_2 + \ldots + x_{k-1} + x_{k+1} = q_k$$

that is,

$$AY = Q$$

where

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 0 & 1 & 1 & \ldots & 1 & 1 \\ 1 & 1 & 0 & 1 & \ldots & 1 & 1 \\ \ldots & & & & & & \\ 1 & 1 & 1 & 1 & \ldots & 0 & 1 \end{bmatrix}$$

and $Y^T = (x_6, x_2, \ldots, x_{k+1})$.

Let X_{k+1} be the column vector with entries all equal to x_{k+1} and let X be the column vector with entries $x_1, x_2, \ldots, x_k$. Then

$$(J - I)X = Q - X_{k+1}$$

where J is the k x k unit matrix and I is the k x k identity matrix.

Then

$$X = (\frac{1}{k-1} J - I)(Q - X_{k+1})$$
$$(\frac{1}{k-1} J - I)Q - \frac{1}{k-1} X_{k+1}$$
$$Q^* - \frac{1}{k-1} X_{k+1}$$

Thus $x_1, x_2, \ldots, x_k$ can be expressed in terms of x_{k+1} as

$$x_1 = q_1^* - \frac{1}{k-1} x_{k+1}$$

$$x_2 = q_2^* - \frac{1}{k-1} x_{k+1}$$

...

$$x_k = q_k^* - \frac{1}{k-1} x_{k+1}$$

Knowing q_i^* we can find the relative order of magnitude of the elements $x_1, x_2, ..., x_k$ (it will be the same as the relative order of magnitude of the corresponding q_i^* values).

Example

Suppose we make three queries about four individuals with the following results

$$x_2 + x_3 + x_4 = 1050$$

$$x_1 + x_3 + x_4 = 70$$

$$x_1 + x_2 + x_4 = 1020$$

Then

$$x_1 = 20 - \tfrac{1}{2} x_4$$

$$x_2 = 1000 - \tfrac{1}{2} x_4$$

$$x_3 = 50 - \tfrac{1}{2} x_4$$

and so $x_1 < x_3 < x_2$.

Note that in some situations a relative compromise with supplementary knowledge could lead to a more serious compromise of the database. Consider for example the situation when it is known that x_i are all nonegative and that any individual with $x_i > 900$ has AIDS while any individual with $x_i < 100$ does not have AIDS.

Then we could infer from the above example that:

$$0 \leq x_4 \leq 40$$

$$0 \leq x_1 \leq 20$$

$$980 \leq x_2 \leq 1000$$

$$30 \leq x_3 \leq 50$$

and hence that the individual corresponding to x_2 must have AIDS while none of the other individuals involved in the queries have AIDS.

Note also that we know rather more than just the order of the X-values, in fact, we know the differences $x_i - x_j$ $(= q_i^* - q_j^*)$ of any two of them. Hence knowing one of the values (if for example the user could plant his/her own X-value into the queries) would result in the complete compromise of all the individuals involved in the queries.

The next theorem shows that relative compromise is possible even if we further restrict the number of queries allowed.

Theorem 2: Let S be a subset of a database, $|S| = k+l$. Then a subset of k individuals of S can be relatively compromised using only k SUM queries with fixed query set size k.

Proof: Construct queries that can be written as the following system of k equations on $k+l$ unknowns.

$$x_2 + x_3 + ... + x_k + x_{k+1} + ... + x_{k+l} = q_1$$

$$x_1 + x_3 + ... + x_k + x_{k+1} + ... + x_{k+l} = q_2$$

...

$$x_1 + x_2 + ... + x_{k-1} + x_{k+1} + ... + x_{k+l} = q_k$$

This can be written as

$$(J - I)X = Q - X_{k+l}$$

where X is the column vector with entries $x_1, x_2, ... x_k$ and X_{k+l} is the column vector with entries all equal to $x_{k+1} + x_{k+2} + ... + x_{k+l}$.

Then

$$X = (\frac{1}{k-1} J - I)(Q - X_{k+l})$$

$$= Q^* - \frac{1}{k-1} X_{k+l}$$

and so

$$x_1 = q_1^* - \frac{1}{k-1}(x_{k+1} + ... + x_{k+l})$$

$$x_2 = q_2^* - \frac{1}{k-1}(x_{k+1} + ... + x_{k+l})$$

...

$$x_k = q_k^* - \frac{1}{k-1}(x_{k+1} + ... + x_{k+l})$$

Since we know the values q_i^* we can find the relative order of magnitude of the elements $x_1, x_2, ..., x_k$.

Example

Consider a set of six individuals and suppose we ask four queries revealing the following SUMs.

$$x_2 + x_3 + x_4 + x_5 + x_6 = 1330$$

$$x_1 + x_3 + x_4 + x_5 + x_6 = 1080$$

$$x_1 + x_2 + x_4 + x_5 + x_6 = 1360$$

$$x_1 + x_2 + x_3 + x_5 + x_6 = 380$$

Then

$$x_1 = \frac{160}{3} - \frac{1}{3}(x_5 + x_6)$$

$$x_2 = \frac{910}{3} - \frac{1}{3}(x_5 + x_6)$$

$$x_3 = \frac{70}{3} - \frac{1}{3}(x_5 + x_6)$$

$$x_4 = \frac{3010}{3} - \frac{1}{3}(x_5 + x_6)$$

and so we deduce that $x_3 < x_1 < x_2 < x_4$.

So far we have considered only the case of relative compromise of queries of fixed query set size k with overlap $k - 1$. Let us now consider a more general situation when the overlap λ is not necessarily equal to $k - 1$. Firstly, we shall consider the case of m linearly independent queries concerning $l (> m)$ individuals, with query set size k and overlap $\lambda \geq l - m$. In this case we can make use of known results (e.g. Davida, 1978) of (complete) compromise using m SUM queries concerning m individuals, with query set size $k - l$ and overlap $\lambda - l$.

Theorem 3: Let S be a subset of a database, $|S| = l$. If $m (< l)$ SUM queries with fixed query set size k^* and over-

lap λ^* lead to the complete compromise of m individuals then a subset of m individuals of S can be relatively compromised using only m SUM queries of query set size $k = k^* + l - m$ with overlap $\lambda = \lambda^* + l - m$.

Proof: (i) $\lambda > l - m$.

Suppose m individuals of S can be compromised using m SUM queries of query set size k^* with overlap λ^*. Let these m queries be written as a system of m linearly independent equations

$$AX = Q$$

where A is an $m \times m$ query matrix, Q is a column vector $(q_1, q_2, ..., q_m)$ and X is a column vector $(x_1, x_2, ..., x_m)$.

Then since A is a nonsingular matrix with constant rowsum k^*, it has an inverse whose rowsum is constant and equal to $\frac{1}{k^*}$.

Now, we can construct queries corresponding to the following equations.

$$AX + Y = Q' \quad (1)$$

where Y is a column vector whose entries are all equal to $x_{m+1} + ... + x_{m+\lambda-\lambda^*}$.

This can be written as

$$A'X' = Q' \quad (2)$$

where $A' = A + B$ and X' is the concatenated vector of X and Y.

Clearly, (2) corresponds to m SUM queries of query set size $k = k^* + \lambda - \lambda^*$ concerning $m + \lambda - \lambda^*$ individuals, with overlap λ.

Now we can solve (1) for $x_1, x_2 ..., x_m$ as

$$X = A^{-1}Q' - \frac{1}{k^*}Y$$

Hence the order of magnitude of $x_1, x_2 ..., x_m$ is the same as the order of magnitude of the entries in $A^{-2}Q'$.

(ii) $\lambda = l - m$.

Then we can use the $m \times m$ identity matrix I in place of A and the proof is essentially the same as for Case (i).

Example

Complete compromise is possible by asking the following seven queries about seven individuals of a database ($\lambda^* = 1, k^* = 3$).

$$x_1 + x_2 + x_3 = q_1$$
$$x_3 + x_4 + x_5 = q_2$$
$$x_1 + x_5 + x_6 = q_3$$
$$x_2 + x_5 + x_7 = q_4$$
$$x_3 + x_6 + x_7 = q_5$$
$$x_1 + x_4 + x_7 = q_6$$
$$x_2 + x_4 + x_6 = q_7$$

We can write this as

$$AX = Q$$

where the query matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

If $q_1 = 500, q_3 = 140, q_4 = 250, q_5 = 270, q_6 = 140, q_7 = 230$ then $x_1 = 100, x_2 = 200, x_4 = 0, x_5 = 10, x_6 = 30, x_7 = 40$.

We shall use the matrix A to form seven queries about nine individuals as follows. Let X be the column vector $(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ and let Y be the column vector with entries all equal to $x_8 + x_9$. Then we can construct the SUM queries corresponding to

$$AX + Y = Q'$$

Then $X = A^{-1}Q' - \frac{1}{3}(x_8 + x_9)$ where

$$A^{-1} = \frac{1}{6}\begin{bmatrix} 2 & -1 & 2 & -1 & -1 & 2 & -1 \\ 2 & -1 & -1 & 2 & -1 & -1 & 2 \\ 2 & 2 & -1 & -1 & 2 & -1 & -1 \\ -1 & 2 & -1 & -1 & -1 & 2 & 2 \\ -1 & 2 & 2 & 2 & -1 & -1 & -1 \\ -1 & -1 & 2 & -1 & 2 & -1 & 2 \\ -1 & -1 & -1 & 2 & 2 & 2 & -1 \end{bmatrix}$$

Suppose the responses to the SUM queries were 550,260,190,300,320,190,280. Then we can express $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ in terms of $x_8 + x_9$ as

$$x_1 = \frac{700}{6} - \frac{1}{3}(x_8 + x_9)$$
$$x_2 = \frac{1300}{6} - \frac{1}{3}(x_8 + x_9)$$
$$x_3 = \frac{1300}{6} - \frac{1}{3}(x_8 + x_9)$$
$$x_4 = \frac{100}{6} - \frac{1}{3}(x_8 + x_9)$$
$$x_5 = \frac{160}{6} - \frac{1}{3}(x_8 + x_9)$$
$$x_6 = \frac{280}{6} - \frac{1}{3}(x_8 + x_9)$$
$$x_7 = \frac{340}{6} - \frac{1}{3}(x_8 + x_9)$$

and so $x_4 < x_5 < x_6 < x_7 < x_1 < x_2 < = x_3$.

Note that the relative compromise of Theorem 3 is only possible for overlap $\lambda \geq l - m$ or, equivalently, for the number of queries $m \geq l - \lambda$. Thus there is a trade off between the number of queries and the overlap.

If we allow m queries about $m + 1$ individuals and m is such that the block design $(m + 1, k, \lambda)$ (Street, 1977) exists then it is possible to get relative compromise given the overlap $\lambda (> 0)$ and fixed query set size K. However, in this case we obtain two disjoint sets of relatively compromised individuals.

Theorem 4: Let S be a subset of a database where $|S| = m + 1$. If the block design $(m + 1, k, \lambda)$ exists then a subset of S can be relatively compromised using m SUM queries with overlap λ and query set size k.

Proof: If m,k and λ are such that the block design$(m+1, k, \lambda)$ exists then we can use for the query matrix the first m rows of the incidence matrix, D, of the design. For the sake of our calculations we assume that the $m+1$st row was also used and that the answer was some unknown constant, say C.

Thus

$$DX = Q$$

where $X^T = (x_1,x_2, ..., x_m,x_{m+1})$ and $Q^T = (q_1,q_2, ..., q_m, C)$, where C is not known. Since D is the incidence of an $(m+1, k,\lambda)$ design it satisfies

$$DJ = kJ$$

$$DD^T = (k - \lambda)I + \lambda J$$

and

$$D^{-1} = \frac{1}{k-\lambda}D^T - \frac{\lambda}{k(k-\lambda)}J$$

Hence

$$X = \left(\frac{1}{k-\lambda}D^T - \frac{\lambda}{k(k-\lambda)}J\right)Q$$

and so $x_1,x_2 ..., x_m, x_{m+1}$ can be expressed in terms of the unknown C,

$$x_i = q_i^* - c_iC$$

where c_i takes on two different values, say c_1 and c_2. Now we can compare all the x_i with the corresponding c_i value equal to c_1 and find their relative order of magnitude. Similarly, we can also order the elements x_i where the corresponding value of c_i is c_2.

Example

Suppose we were allowed to make the 6 SUM queries about 7 individuals which correspond to the following set of 6 linearly independent equations in 7 unknowns

$$x_1 + x_3 + x_7 = 1300$$
$$x_1 + x_2 + x_4 = 230$$
$$x_2 + x_3 + x_5 = 1010$$
$$x_3 + x_4 + x_6 = 1030$$
$$x_4 + x_5 + x_7 = 120$$
$$x_1 + x_5 + x_6 = 210$$

The query matrix of these 6 queries corresponds to the first 6 rows of the 7 x 7 incidence matrix D of the (7,3,1) block design. Suppose we asked the query corresponding to the last row of the matrix D and the system refused to answer that query. That is, we have

$$x_2 + x_6 + x_7 = C$$

where C is unknown. Then the system of equations can be written as

$$DX = Q$$

where

$X^T = (x_1,x_2,x_3,x_4,x_5,x_6,x_7)$, $Q^T = (1300,230,1010, 1030,120,210,C)$ and

$$D = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The inverse of D is

$$D^{-1} = \frac{1}{6}\begin{bmatrix} 2 & 2 & -1 & -1 & -1 & 2 & -1 \\ -1 & 2 & 2 & -1 & -1 & -1 & 2 \\ 2 & -1 & 2 & 2 & -1 & -1 & -1 \\ -1 & 2 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & -1 & 2 & 2 & -1 \\ -1 & -1 & -1 & 2 & -1 & 2 & 2 \\ 2 & -1 & -1 & -1 & 2 & -1 & 2 \end{bmatrix}$$

$$X = D^{-1}Q = \frac{1}{6}\begin{bmatrix} 1320 - C \\ -180 + 2C \\ 6120 - C \\ 240 - C \\ 120 - C \\ -180 + 2C \\ 360 + 2C \end{bmatrix}$$

Thus we can calculate the values of x_1,x_2,x_3,x_4,x_5,x_6 and x_7 in terms of C and since C appears with only two different coefficients we can find the relative order of magnitude of two subsets of the 7 unknowns, namely

$$x_5 < x_4 < x_1 < x_3 \text{ and } x_2 = x_6 < x_7.$$

Note that if in the above example the sixth query were also disallowed so that the result of that query was unknown, say C', then we could still get some relative compromise. In particular, we could compare x_1 with x_5 ; x_7 and x_3 with x_4. This is easily seen by inspecting the last two columns of the inverse matrix D^{-1}.

In general, using the incidence matrix of a block design $(m+1,k,\lambda)$ for a query matrix we can guarantee to get some relative compromise whenever $m+1-n$ queries about $m+1$ individuals (with constant query set size k and overlap λ) are allowed provided $m+1 > 2^n$.

CONCLUSION

In this paper we have described the idea of relative compromise of statistical databases and some conditions under which it can occur using SUM type of queries. We also showed that in some cases relative compromise with supplementary knowledge can lead to a more serious compromise of a database. It remains an open problem to find the general conditions (in terms of the number of records in a database, query set size, query overlap and the minimum number of queries needed) under which relative compromise can occur.

ACKNOWLEDGEMENT

The advice and suggestions by the referees has been much appreciated by the authors.

REFERENCES

DAVIDA, G.I., LINTON, D.J., SZELAG, C.R. and WELLS, D.L. (1978): "Database security", *IEEE Transactions on Software Engineering*, Vol. SE-4, No. 6, pp. 531-533.

DENNING, D.E.R. (1982): *Cryptography and Data Security*, Addison-Wesley, Sydney.

DOBKIN, D., JONES, A.K. and LIPTON, R.J. (1979): "Secure databases: Protection against user influence", *ACM Transactions on Database Systems*, Vol. 4, No. 1, pp. 97-105.

FERNANDEZ, E.B., SUMMERS, R.C. and WOOD, C. (1981): *Database Security and Integrity*, Addison-Wesley.

SEBERRY, J. and PIEPRZYK, J. (1988): *Cryptography: An Introduction to Computer Security*, Prentice-Hall, Sydney.

STREET, A.P. and WALLIS, W.D. (1977): *Combinatorial Theory: An Introduction*, Charles Babbage.

BIOGRAPHICAL NOTES

Jennifer Seberry is Professor and Head of the Computer Science Department at University College, ADFA, Canberra. Dr Seberry is especially interested in Cryptography, Authentication and Computer Security. She has co-authored six books including Cryptography: An Introduciton to Computer Security *with Josef Pieprzyk (Prentice Hall, 1988) and* The Cryptographic Significance of the Knapsack Problem *with Luke O'Connor (Aegean Press, 1988). Professor Seberry is a member of the Australian Computer Society and the Australian Operations Research Society since 1970. She is also a member of the International Association for Cryptologic Research.*

Mirka Miller is a lecturer in the Department of Mathematics, Statistics and Computing Science, University of New England, Armidale. Her main interests in computer science are databases, especially security and integrity of databases. She is a member of the Australian Computer Society, the European Association for Theoretical Computer Science and the Australian Mathematical Society. Ms Miller is currently working on her PhD thesis in security of databases, under the supervision of Professor Seberry.